EDUCAUSE

ANNUAL CONFERENCE 2017

Measuring ROI in Cyberinfrastructure

PRESENTED BY:

Craig A. Stewart • Matthew R. Link • Julie Wernert

PERVASIVE TECHNOLOGY INSTITUTE

#EDU17

Goal and outline

Goal: Discuss concepts and techniques you can take home and use to explain the value of investment in research cyberinfrastructure

Outline:

- Value added
- Knowledge creation
- Cost avoidance
- Estimating regional impact
- ROI at the campus level
- XDMoD Value Analytics – an open source tool you can use to assess ROI of your research cyberinfrastructure to your campus and institution
- A very few comments about research cyberinfrastructure nationally

Working definition of cyberinfrastructure – focused on research

"Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible."

From Stewart, C.A., S. Simms, B. Plale, M. Link, D. Hancock and G. Fox. What is Cyberinfrastructure? In: Proceedings of SIGUCCS 2010. (Norfolk, VA, 24-27 Oct, 2010). http://portal.acm.org/citation.cfm?doid=1878335.1878347

Turning to finances - some definitions

- **Value Added:** "an activity that increases the worth of the product or services to the customer"
- **Cost avoidance:** "the practice of finding acceptable alternatives … and/or not spending money for unnecessary goods or services." This is measured as the cost difference between doing something one way vs. a hypothetical other way.
- **Return On Investment (ROI):** "a ratio that relates income generated…to the resources (or asset base) used to produce that income," calculated typically as "income or some other measure of return on investment." Values > 1.0 indicate that return > investment.

From Kinney, M.R., and C.A. Raiborn. 2013. Cost accounting foundations and evolutions. South-Western, Mason, OH. 832 pp.

Value added and ROI - Societal benefits

Year of study	Subject	Annual rate of return (%)
1979	Tomato harvester	37–46
1968	Agricultural research	35–40
1968	Agricultural research	28–47
1979	Agricultural research	37
1979	Agricultural research	45
1981	Agricultural research	37
1958	Hybrid corn	20–40
1991	All academic science research	28
1993	Agricultural research	43–67
2000	Pharmaceuticals	30+
1967	Poultry	21–25

Rising Above the Gathering Storm (National Academies, 2006); Scott, G. et al. The Economic Returns of Basic Research and the Benefits of University–Industry Relationships Science and Technology Policy Research (Univ. Sussex, 2001).

Value added - Apon et al. – impact of investment in high performance computing on knowledge creation

- Investment in local High Performance Computing (HPC) facilities increases grant income and publication rate
- Local HPC facilities speed innovation in many but not all disciplines

Apon, A.W., S. Ahalt, V. Dantuluri, C. Gurdgiev, M. Limayem, L. Ngo and M. Stealey. 2010. High Performance Computing Instrumentation and Research Productivity in U.S. Universities. Journal of Information Technology Impact, 10(2), 87-98.

Apon, A.W., L.B. Ngo, M.E. Payne and P.W. Wilson. 2014. Assessing the effect of high performance computing capabilities on academic research output. Empirical Economics 48: 283-312. DOI: 10.1007/s00181-014-0833-7

Cost avoidance on required class texts – an unequivocal case

- Cost avoidance as Δ between actual and list costs of physical texts to date: $21,673,338
- IU reports half of this, figuring that some students don't buy texts, buy used, or sell books at semester's end

	2012	2013	2014	2015	2016
Course Sections	328	679	1,166	1,714	2,279
Adoptions*	690	1,167	1,751	1,912	2,590
Students	12,251	24,290	32,923	48,814	67,327

**"Adoptions" refers to a single course item (e.g., digital book, and a section may have one or more).*

From: https://er.educause.edu/articles/2017/10/student-engagement-with-etexts-what-the-data-tell-us

Cost avoidance – licensing research software

- The IU Pervasive Technology Institute licenses a number of research software packages for use by the IU community that are distributed without cost to the IU community
- Equivocal because it's clear that not everyone that downloads the software would have purchased it at academic list, but probably half of the users would have purchased the software at academic list prices

Year	Actual expenditures	Cost if purchased at academic list price	Cost avoidance
FY 2017	$434,568	$2,441,000	$2,006,432

Cost avoidance – IU local clusters vs. AWS

- One year period July 1, 2016 – June 30, 2017
- Karst specs: Intel 2650 v2 (Ivy Bridge) with ≥ 32 GB RAM
- AWS specs: c4.4xlarge 16 vCPU with 30 GB RAM (hyper-threaded Intel Haswell core)
- Used partial upfront price for reserved AWS costs
- Karst hardware cost amortized over 5 years, includes sysadmin time, and direct/indirect facilities costs
- 25,003,016 core hrs consumed on Karst equivalent to 1,562,689 instance hrs

System	On Demand	1-year reserved	3-year reserved
c4.4xlarge	$1,243,900.05	$754,778.55	$515,687.21
Karst	$414,571.24	$414,571.24	$414,571.24
1-year Cost Avoidance	**$829,328.81**	**$340,207.31**	**$101,115.97**

Cost avoidance - HEPCloud

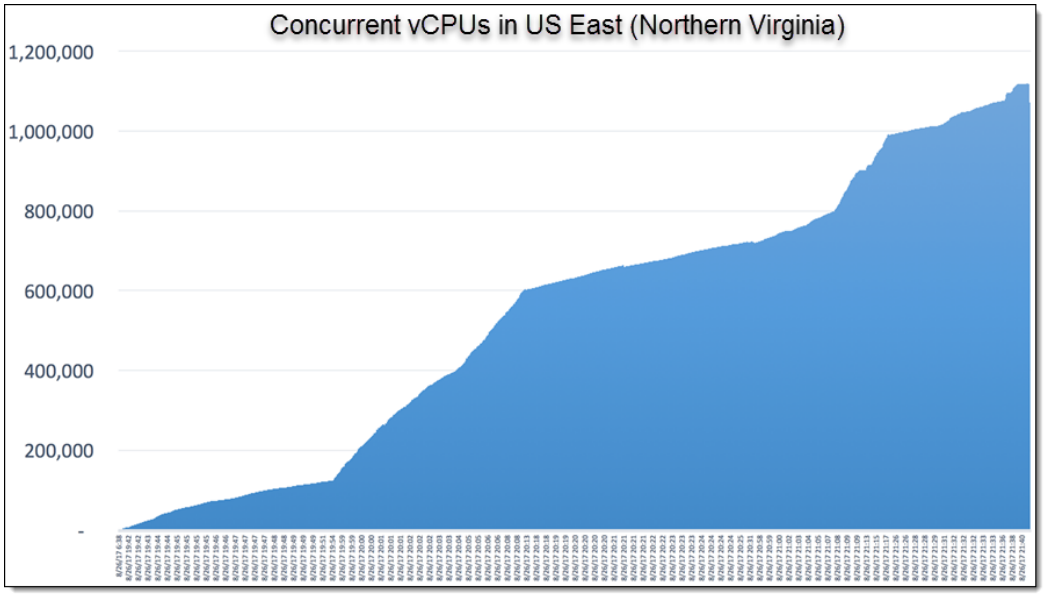
- CMS Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at CERN in Geneva, Switzerland
- Purchased AWS services on the “spot market”
- $211,985 spent on AWS services during trial
- Cost comparison
  - Fermilab CMS Tier-1 $0.009 ± 25%
  - AWS $0.014 ± 12%
- Functionality comparison:
  - Not all workloads could be run on AWS
  - AWS provides capacity elasticity that Fermilab resources do not provide

https://link.springer.com/article/10.1007/s41781-017-0001-9

HPC Cluster with more than 1M vCPUs on Amazon EC2 Spot Instances

using US East (Northern Virginia) region

Issues related to “describe-instance”, API limits, etc., prevented a faster ramp-up. These were quickly resolved during the study.



From Professor Amy Apon, Clemson Univ. Used by permission. May not be reused without permission from Dr. Apon. Supported in part by NSF Grant #0722625.

EDUCAUSE ANNUAL CONFERENCE 2017

#EDU17

Contact: Amy Apon, aapon@Clemson.edu

Runtime and Cost Summary

Number of vCPUs at peak: 1,119,196
Total vCPU hours used: 1,832,923
Total Spot Cost: $31,559
Average vCPU hour cost: $0.0172

Per-second billing was used

Typical instance setup time is 1-2 minutes (performed concurrently across instances)

No costs are incurred after execution completes

Contact: Amy Apon, aapon@Clemson.edu

 Supported in part by NSF Grant #0722625.

Regional financial impact multipliers and other methods of analyzing impact

- Recent study from University of Illinois Urbana-Champaign
- Estimated impact of Blue Waters
- Used “Regional Impact Multipliers” to estimate total economic impact of Blue Waters supercomputer and other methods
    - Oct 2007-June 2016 Blue Waters generated $634 M direct and indirect impact on economy of Illinois compared to $60.3M State of Illinois investment. (Based on IMPLAN RIM methodology)
    - Led to creation of 3,439 direct and indirect FTE jobs

http://www.ncsa.illinois.edu/news/story/ncsas_blue_waters_project_provide

Return on investment at the campus level

- Higher education is primarily about creating and disseminating knowledge. Much of what we do is difficult or impossible to evaluate in dollar and sense terms.
- But CFOs by necessity think in terms of dollars and cents
- So where do we see research cyberinfrastructure affecting campus budgets?
    - In teaching… sure, but in ways that are hard to relate to tuition income
    - In research ... absolutely, and while there are difficulties in relating CI investments to research-oriented income to a campus, it is a tractable problem

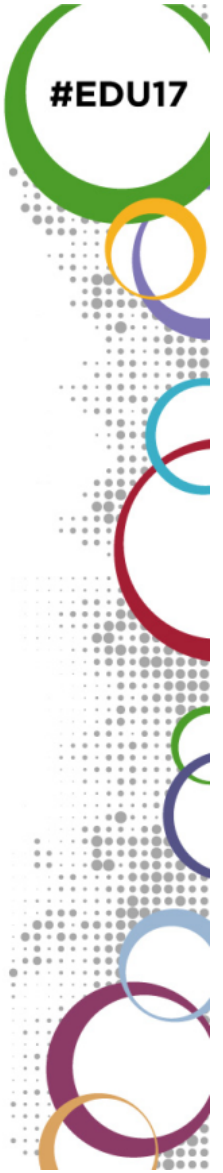
What's sorts of grant-related income that matters most to campus leaders?

Academic leaders

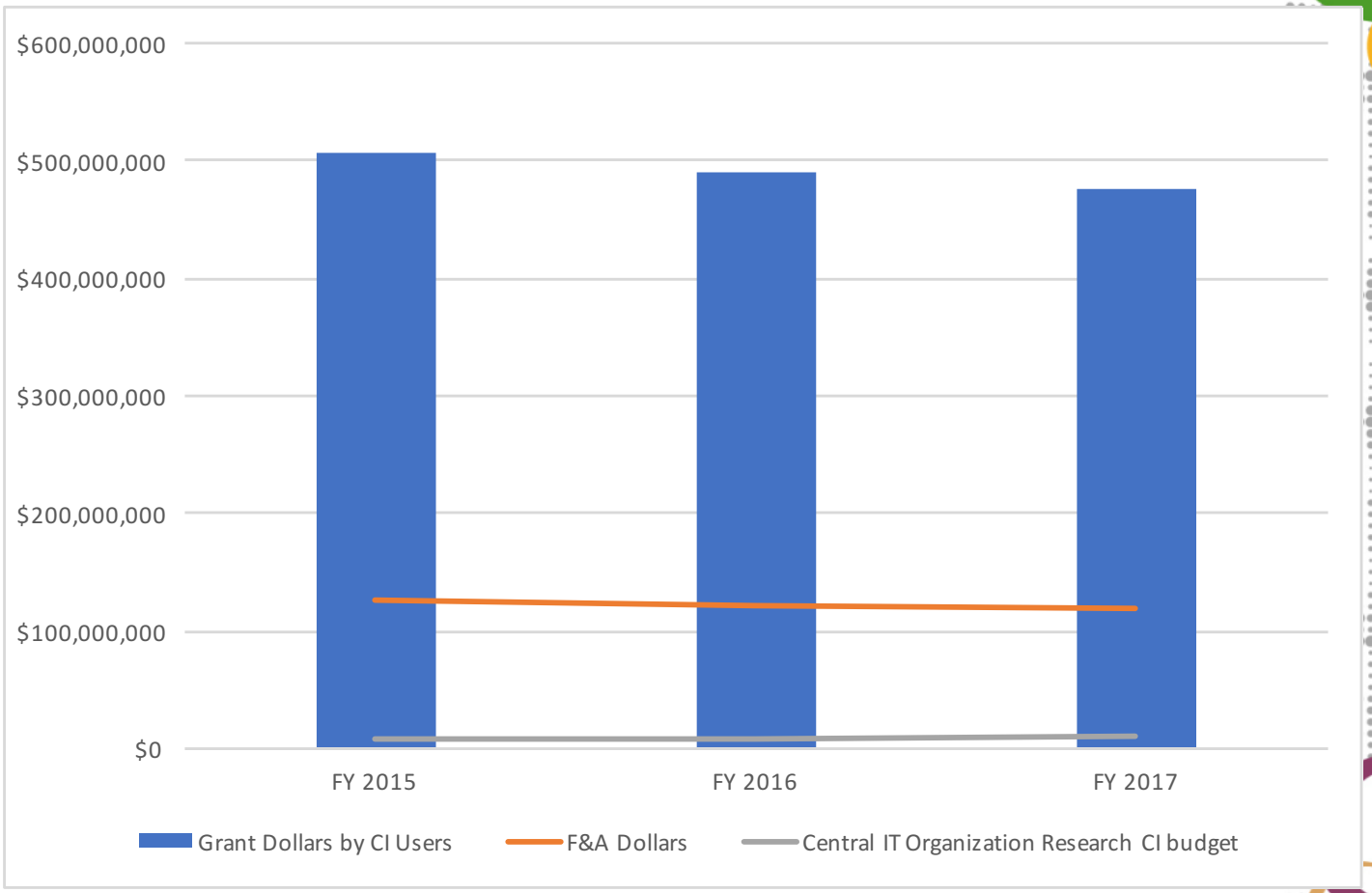
- Publications
- Impact metrics of publications
- Tech transfer
- Nobel and similar prizes

Financial leaders

- Grant direct costs - sure
- Facilities and Administration $s



IU analysis of relationship between CI usage, grant $s, F&A $s, and cost of CI



Can we yet state that the ROI on IU's investment in cyberinfrastructure is > 1?

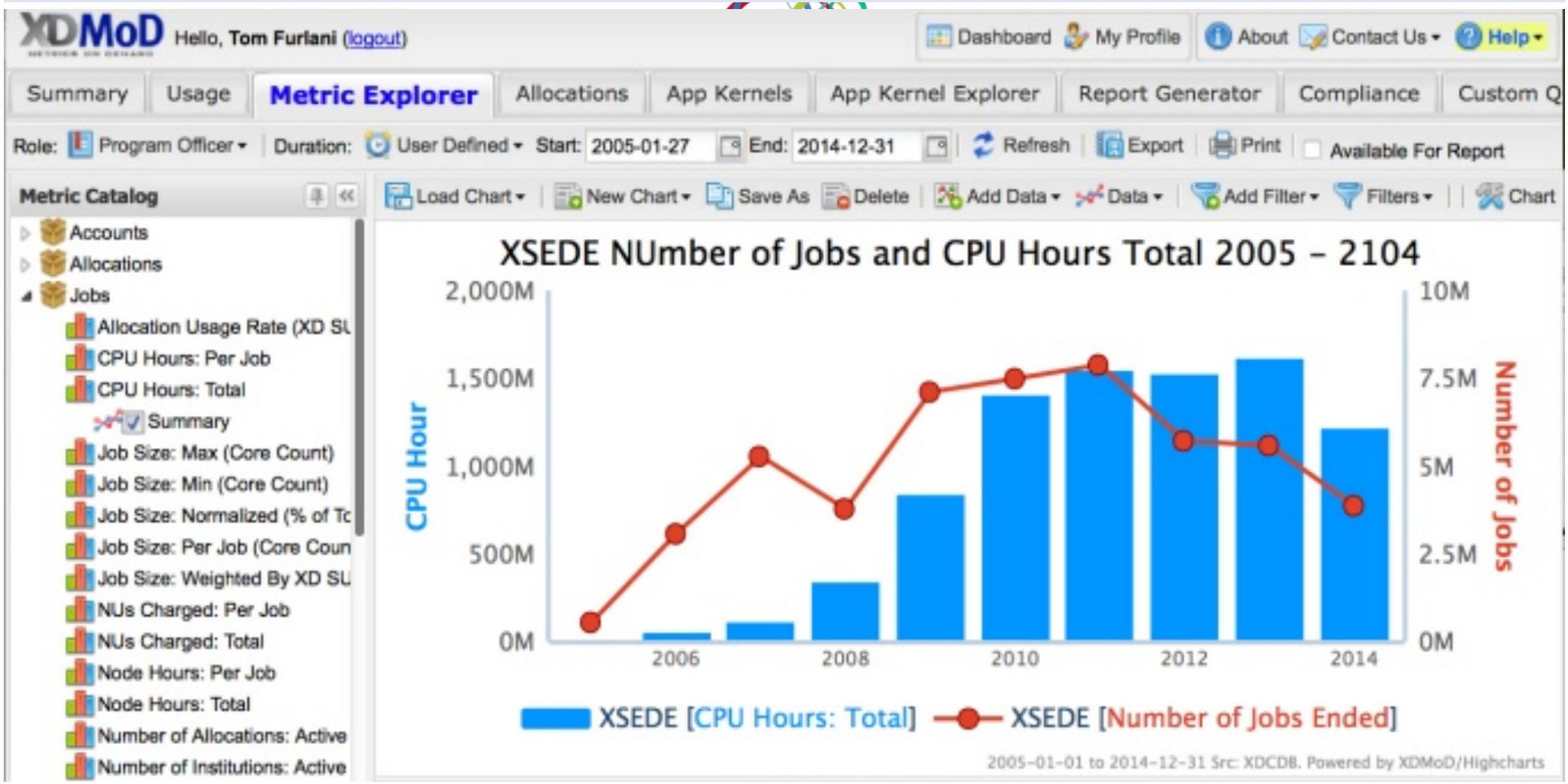
- Not quite yet
- Next step: survey PIs and ask how important CI was to their success in receiving each particular grant award
- Calculate F&A income that is directly ties to the IU Pervasive Technology Institute and see if it is > 1

Open XDMoD Value Analytics – lets you do this analysis for your institution

Open XDMoD

- Developed by the University at Buffalo Center for Computational Research
- Comprehensive resource management for HPC systems
- Provide detailed operational and usage data
- Support optimization of HPC resource utilization
- Facilitate planning and analysis

VA module will be integrated into local[1] version



Open XDMoD Value Analytics

- Enables academic institutions to better understand Return On Investment (ROI) on advanced Cyberinfrastructure (CI)
- Shows the value of
    - Fostering collaboration
    - Supporting scientific publications
    - Show relationship between campus CI and external grant funding
- Local XDMoD installation required – all data kept locally

Data capabilities

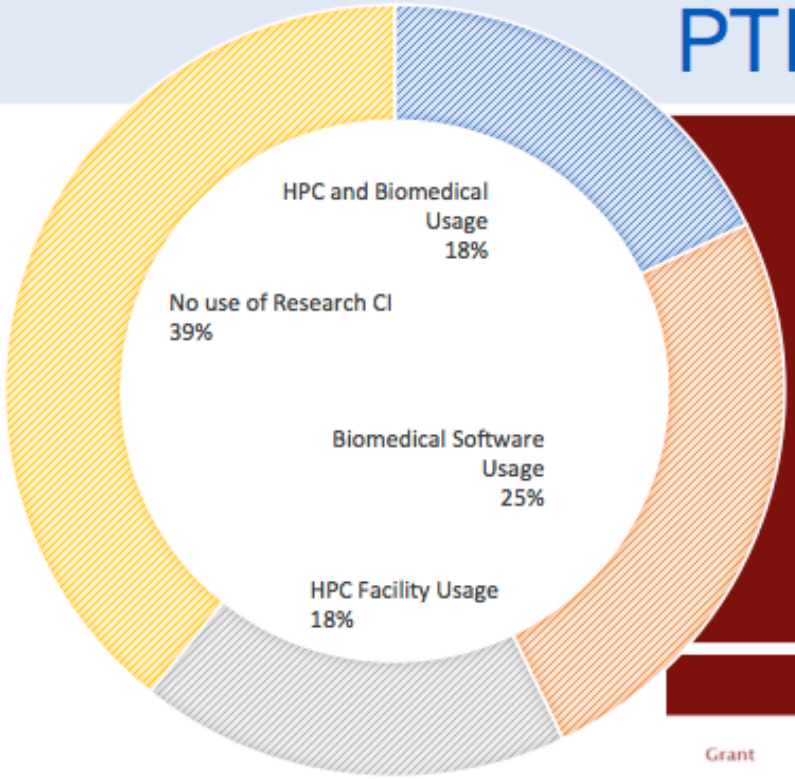
If your local financial management organization will let you retrieve data from its data about grant incomes:

- Now: Kuali Financial Management System export -> JSON convert -> XDMoD_VA ingest
- Future: Add SCT Banner
- With more funding, could add other FMSs

If your local financial management organization will NOT let you retrieve data from its data about grant incomes:

- Scripts to import NSF and NIH grant data from the NSF and NIH web sites

PTI: Funding analytics



Screen shots of PTI developed statistics tool.

RT Stats

Biochemistry/Molecular Biology (Indianapolis)

School Of Medicine

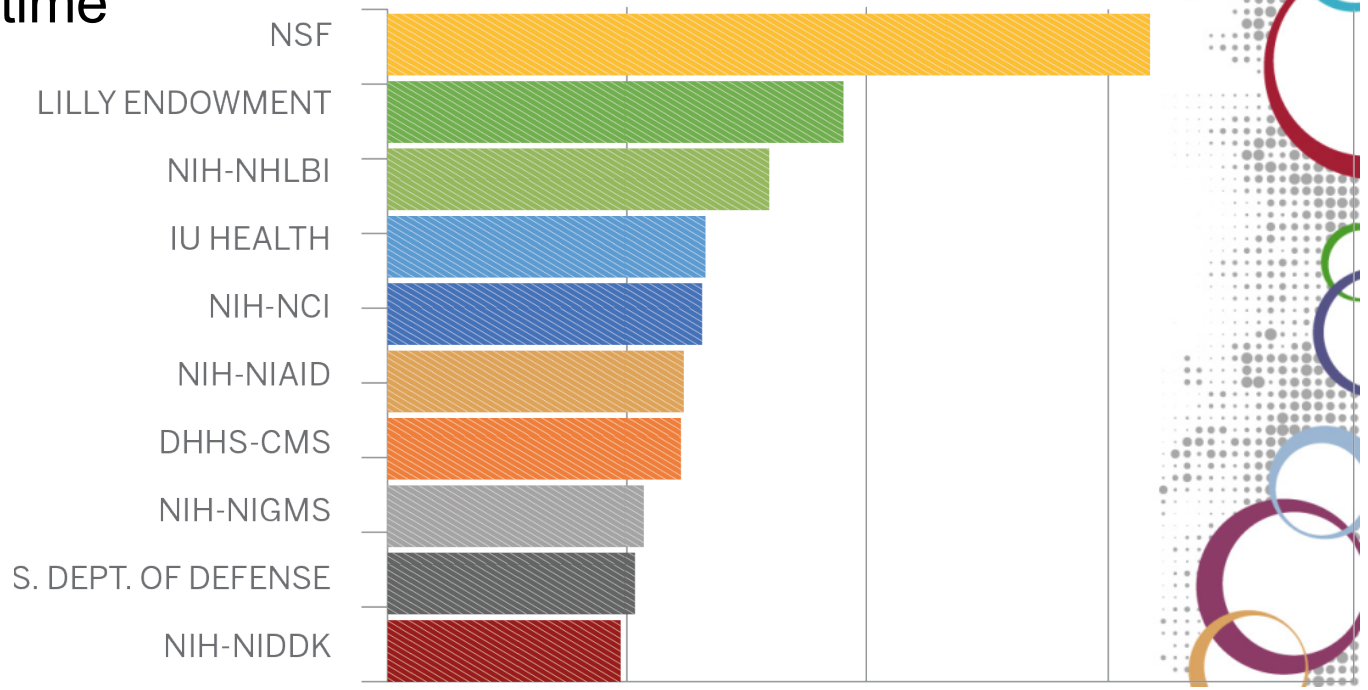
Biochem Auxiliary Services	Bioinformatics	Biom Chemical Genomics Program	Diabetes Basic Science Res Ctr
Abdul Sater, Zahi A	Arthur, Jack W	Cerabona, Donna	Conteh, Abass
Craven, Kelly	Edenberg, Howard J	Folck, Anthony F	Fox, Melanie J
Fusakio, Michael	Gendron, Jaimie M	Georgiadis, Millie M	Heyen, Joshua W
Heyerdahl, Darcy	Hoang, Quyen Q	Huang, Fei	Hunter, Gerald O

Grants
24 totaling $6,056,352.00

Grant	Project	Status	Dates	Award Total
043956-00005B	Docking uPAR for Selective Targeting of Cancer Metastasis	New	January 2015 - June 2016	$180,000.00
044265-00004B	The Bipolar Genome Study	New	May 2014 - April 2016	$87,728.00
053198-00014B	Collaborative Study on the Genetics of Alcoholism (COGA)	New	September 2015 - August 2016	$1,480,906.00
054541-00005B	Regulation of RNA Polymerase II Transcription by the Phosphatase Rtr1	New	August 2015 - July 2016	$291,461.00
054541-00007B	Regulation of RNA Polymerase II transcription by the phosphatase Rtr1	New	August 2015 - July 2016	$34,710.00
056553-00006B	Non-Homologous End Joining Repair in Human	New	April 2015 - March 2016	$319,550.00
056564-00005B	Metabolic stress resones and EIF2 kinase GCN2	New	May 2014 - April 2016	$343,929.00
056564-00008B	Metabolic stress responses and eIF2 kinase GCN2	New	May 2014 - May 2016	$116,099.00
057512-00006B	Early binge drinking and gene regulation	New	September 2015 - August 2016	$225,424.00

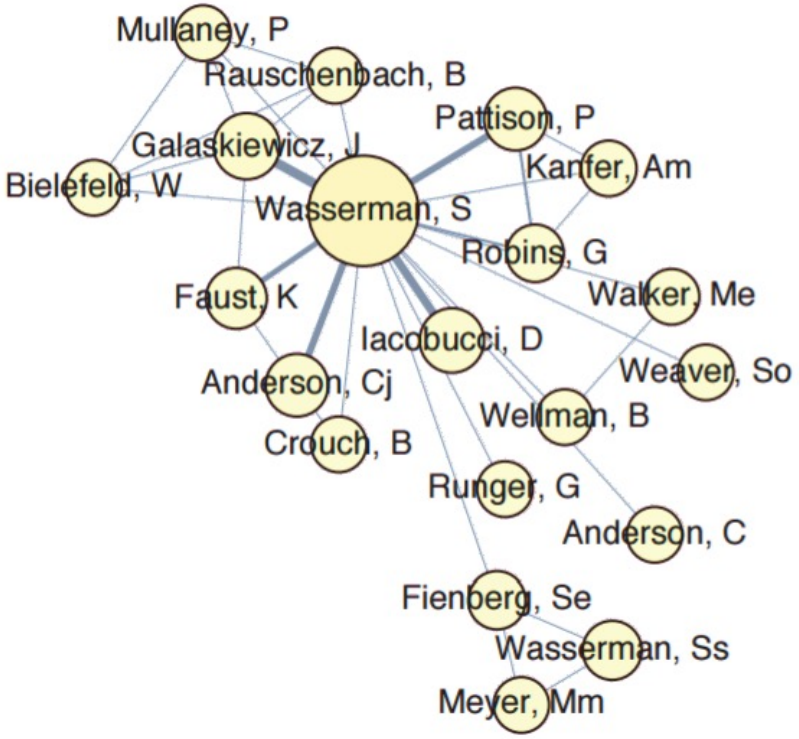
Financial analysis capabilities

- Grant $s by funding agency
- Grant income by organizational subunit
- Drill down to individual investigators
- Look at grants over time



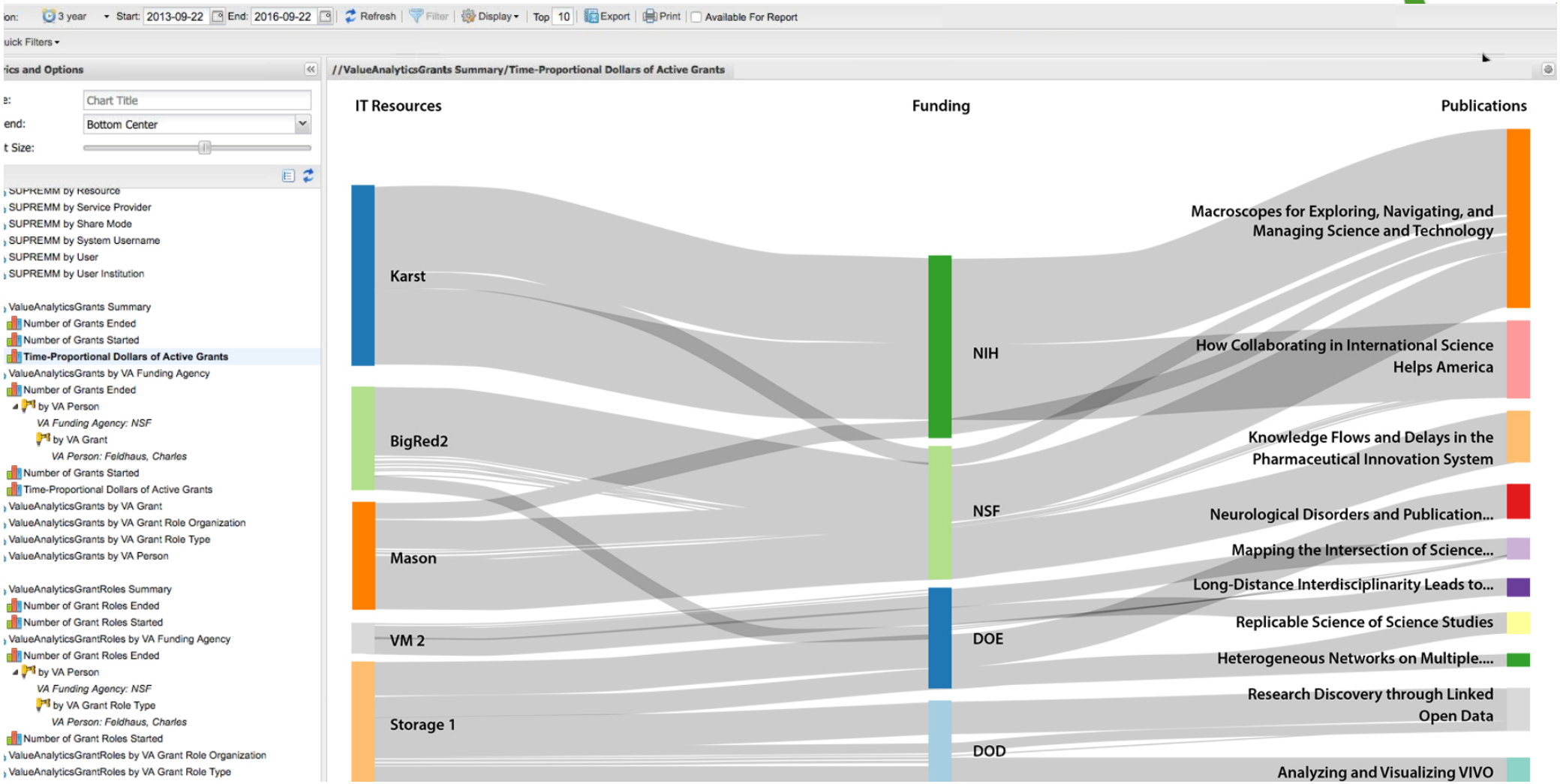
Publication analysis capabilities

- Access to NIH grant and publication data available via NIH Exporter (https://exporter.nih.gov/)
- Working to develop schema for more generalized ingest (from sources such as NSF, Scopus)
- Understand collaboration within and beyond your institution with co-authorship networks



Everything all at once

#EDU17



Roadmap

Early 2018| Final release

July 2017 | Beta release

June 2017 | Cross-referencing grants and HPC usage data

Spring 2017 |

Beta tester signup
Publication data available.
Aggregation charts available.

February 2017

Pre-beta testing under way
Grant and PI information available in XDMoD.

Nobel prizes

- No one is quite sure what it's worth to a university when a faculty member wins a Nobel prize… but it's a lot
- Nobel prizes are a trailing indicator, and the awardee selection process is not perfect
- But if you can tie your campus cyberinfrastructure to a Nobel prize awarded to one of your faculty members, that's great.
- Two national CI services can do so: XSEDE (Extreme Science and Engineering Discovery Environment) and OSG (Open Science Grid) can claim 3 and 2, respectively
- IU thinks we have maybe one good possibility...

At the end of the day*….

- Scientific benefits (new knowledge)
  - Functionality Benefits (things you can do thanks to local CI)
  - Cloud computing has a role in functionality but right now most analyses are that local research CI is less expensive than commercial clouds
- Workforce development benefits
- Financial Cost
- Financial Benefits
- But local facilities are CERTAINLY NOT a vanity project (cf. Chronicle of Higher Education http://www.chronicle.com/article/Supercomputers-a-Status/241564?key=RZCWdyPeVKgP6T2Rw68ZEHKrcSTCFCab8HokInkO4gjWjawlu6km6Pw8T72jvT-Vc1VCTW8tM19lLVJ3Ymp4Mjl1Q0g0V2p2UHNyNzhpa2lmRDFuSjExcGhKbw

EDUCAUSE ANNUAL CONFERENCE 2017

*This slide inspired in part by email commentary from Alan Sill (alan.sill@ttu.edu)

#EDU17

As you talk with your own financial leaders

- **Committed fixed costs:** "a cost related to either the long-term investment in plant and equipment of a business or the organizational personnel who are deemed essential…."
    - Your local HPC system makes computing cost largely a committed fixed cost, not a variable operating cost
- **Cost Center:** "A responsibility center in which the manager has the authority to incur costs and is evaluated on the basis of how well costs are controlled"
- **Profit Center:** "a responsibility center for which the manager is accountable for generating revenues and planning and controlling all expenses"
- *Think about making the argument locally that you represent a profit center rather than cost center*

Definitions from Kinney, M.R., and C.A. Raiborn. 2013. Cost accounting foundations and evolutions. South-Western, Mason, OH. 832 pp.

Additional information and thanks

Interesting articles:

- McIlwain, C. (2010). What science is really worth. Nature, 465(10), 682–684.
- Preuss, M. (2016). Return on Investment and grants. Research Management Review, Volume 21, Number 1 (2016). http://www.ncura.edu/Portals/0/Docs/RMR/2016/v21_n1_Preuss.pdf
- Stewart, C.A., Roskies, R., Knepper, R., Moore, R.L., Whitt, J., & Cockerill, T.M. (2015) XSEDE Value Added, Cost Avoidance, and Return on Investment. Proceedings of the 2015 XSEDE Conference: Scientific Advancement Enabled by Enhanced Cyberinfrastructure, doi: http://dx.doi.org/10.1145.2792745.2792768.

Acknowledgements:

- This work has been supported in part by NSF award 1566393 (EAGER: Open XD Metrics on Demand Value Analytics, Matthew R. Link, PI)
- This work has been supported by the Indiana University Pervasive Technology Institute (pti.iu.edu). PTI has been aided by major grants from the Lilly Endowment.
- Opinions expressed here are those of the speaker and do not necessarily represent the opinions of any funding agencies.
- Thanks to PTI staff who have contributed to this talk, particularly Abhinav Thota, David Y. Hancock, and Winona Snapp-Childs.

EDUCAUSE
ANNUAL CONFERENCE 2017

PERVASIVE TECHNOLOGY INSTITUTE

#EDU17

License terms

- Please cite as: Stewart, C.A., M.R. Link, J. Wernert. 2017. Measuring ROI in Cyberinfrastructure at the Campus and National Level. Presentation. Presented EDUCAUSE Annual Conference 31 Oct – 3 Nov 2017, Philadelphia, PA. https://doi.org/10.5967/K85Q4T8S
-

Questions and answers

…. And thank you for your kind attention